UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, STATION A, MONTRÉAL, QUÉBEC, H3C 3J7

A MODAL CHOICE FUNCTION FOR WINNIPEG
- CALIBRATION AND VALIDATION

by

R. Cléroux, M. Florian and S. Galarneau

This research was sponsored in part by a contract with Transport Canada and by a grant from the National Research Council.

Université de Montréal

Centre de recherche sur les transports - Publication # 135
Département d'informatique et de recherche opérationnelle
- Publication # 320

Avril 1979

Dépôt légal - 2e trimestre 1979

ABSTRACT

We describe in detail the calibration of a modal choice function for the City of Winnipeg, Manitoba. The data used for this work was obtained from a survey performed by the City of Winnipeg in the spring of 1976 and from models of the road and transit networks. Some validation tests are reported. This function is being used currently by the City of Winnipeg for transportation planning purposes.

Introduction

We report in this paper on the calibration of a modal split function for the City of Winnipeg, Manitoba. A modal split function is used in the context of transportation planning and its aim is to predict the relative shares of the main modes used for travel, usually during the peak hours. Disaggregate type modal split functions are calibrated by using data about individual trip makers and predict the probability that an individual will choose a particular mode for his trip; share predictions require the use of some appropriate aggregation procedure. Aggregate type modal split functions are calibrated by using data which is aggregated to a zone level; for transportation planning purposes the urban area is usually subdivided into zones that are intended to have homogeneous socio-economic characteristics. Such modal split functions predict directly the shares of each mode of travel.

The work on which we report represents part of a larger project [3] that consisted of the implementation of a new urban transportation planning method that simultaneously predicts the mode choice and route choice on the road and transit networks. Roughly speaking, the model underlying this method [2] combines a modal split function with an equilibrium type [6] car trip assignment and a transit assignment model. However, the modal split function is one of the most vital components of this model since it must be used for prediction in conjunction with the road and transit network models. Therefore its calibration required the measurement of travel times on the road and transit network models.

The modal split function which we calibrated is an aggregate logit type function. For the purpose of the calibration we obtained the auto travel times from the road network assignment and the transit travel times (or impedances) from the transit network; in addition, we had zonal average values for several socio-economic variables which are described in detail below.

The paper is organized in the following way: first we describe the source of the data used, then the construction of the travel time and socio-economic variables used and then we present the details of the calibration and some validation tests for the modal split function obtained.

The Origin/Destination Survey

For the purpose of transportation planning the City of Winnipeg is subdivided into 147 zones.

The basic data which was used for the calibration of the modal split function are the results of the O/D survey that was carried out by the City of Winnipeg in the spring and summer of 1976. The survey was carried out in large part by home interviews of a sample of 20% of households. The actual sample size obtained was roughly 17%, after refusals and rejects had been taken into account. In addition, a survey questionnaire to be returned by mail was distributed to the students of the three Winnipeg Universities; the effective sample of student trips was approximately 23%. The total sample, consisting of the individual detailed data, amounted to

52,424 individual questionnaires and these data were transmitted to us on a magnetic tape by the City of Winnipeg. Then, a subsample selection procedure was applied in order to separate the trips which occur during the 7:30 - 8:30 AM peak from all the trips that are taken over a 24-hour period. There were 17,761 individual records in the peak hour sub-sample. The selection procedure is described in detail in [4].

The Impedance (Travel Time) Variables

The number of trips which occurred during the peak hour were expanded by the proportion of the sample in each zone, (that we obtained from the City of Winnipeg) in order to obtain the following Origin/Destination matrices:

- auto drivers and passengers (1), auto drivers (1')
- transit passengers (2)
- total trips (3) = ((1) + (2))
- modal split (4) = ((1) / (3))
- car occupancy (5) = ((1) / (1'))

A summary by origin and destination of O/D Matrices (1) and (2) are given in Figures 1, 2, 3, 4, while an average of (4) by origin and destination is given in Figures 5 and 6. (The auto drivers and passengers O/D matrix shall be refered to as the auto O/D matrix.) The auto O/D matrix was scaled by the appropriate factor to account for other than work trips to obtain the total auto O/D matrix and this last was assigned to the road network by using an equilibrium traffic assignment method. A by product

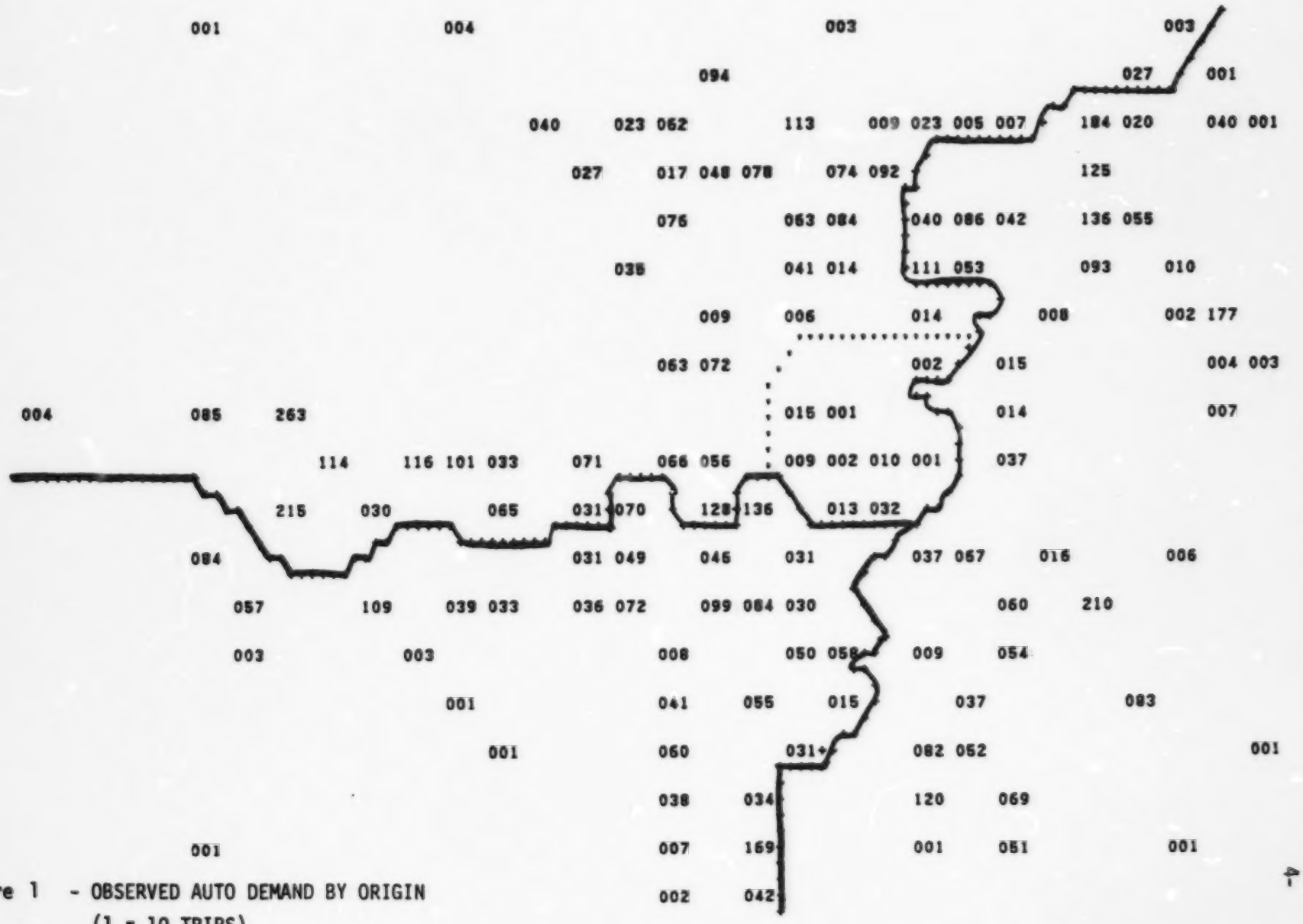

Figure 1 - OBSERVED AUTO DEMAND BY ORIGIN
(1 = 10 TRIPS)

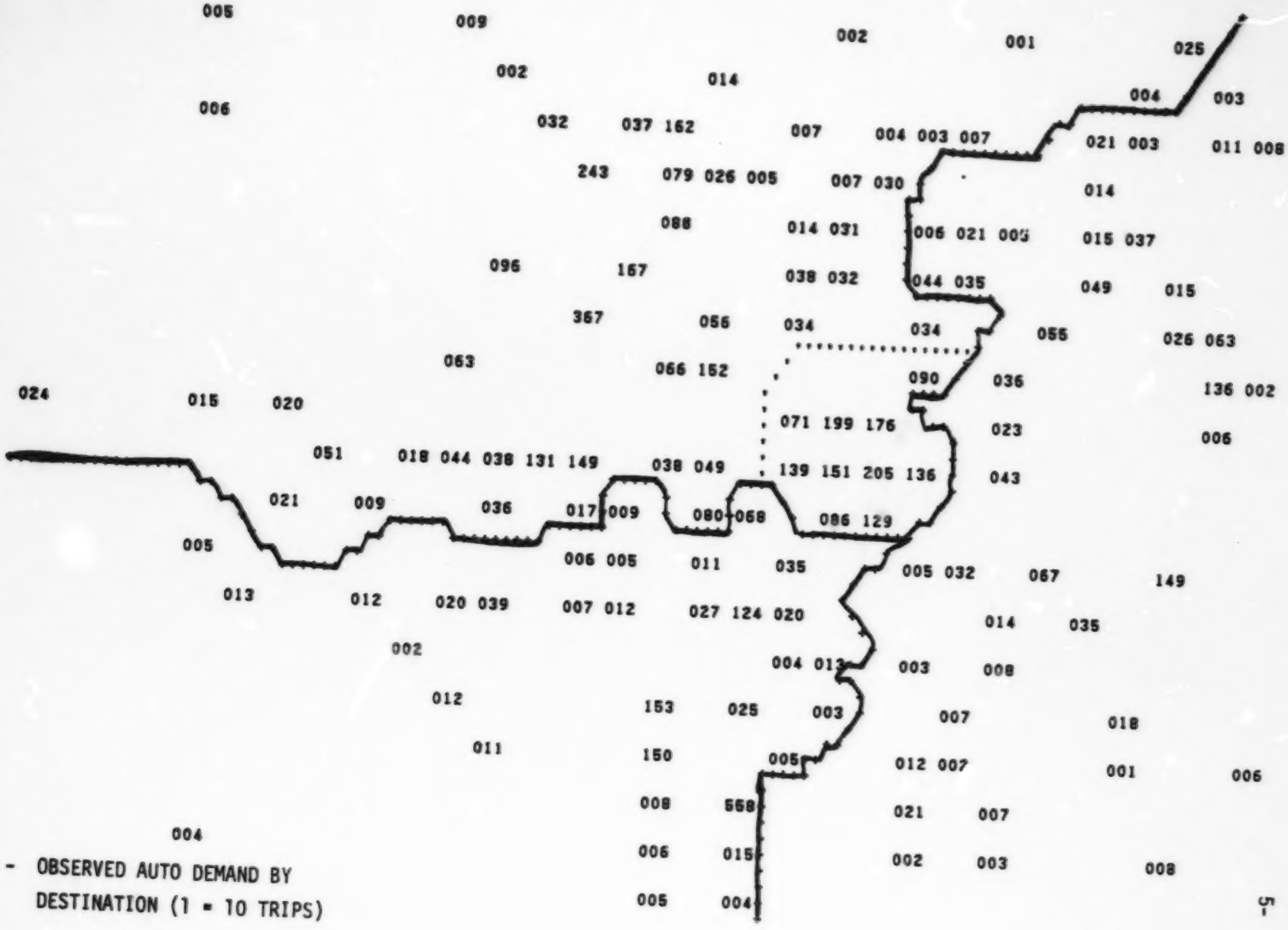

Figure 2 - OBSERVED AUTO DEMAND BY DESTINATION (1 = 10 TRIPS)

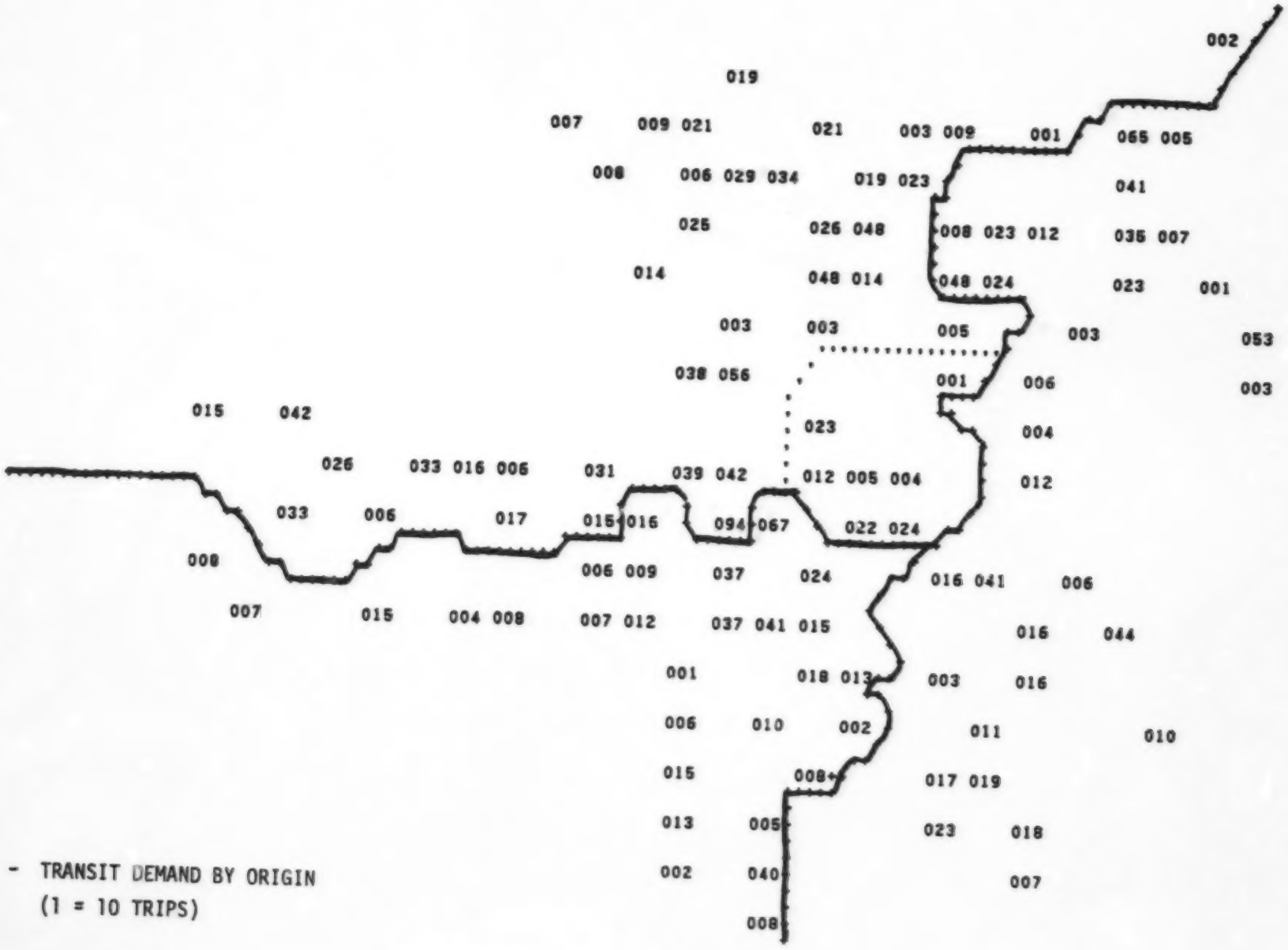

Figure 3 - TRANSIT DEMAND BY ORIGIN
(1 = 10 TRIPS)

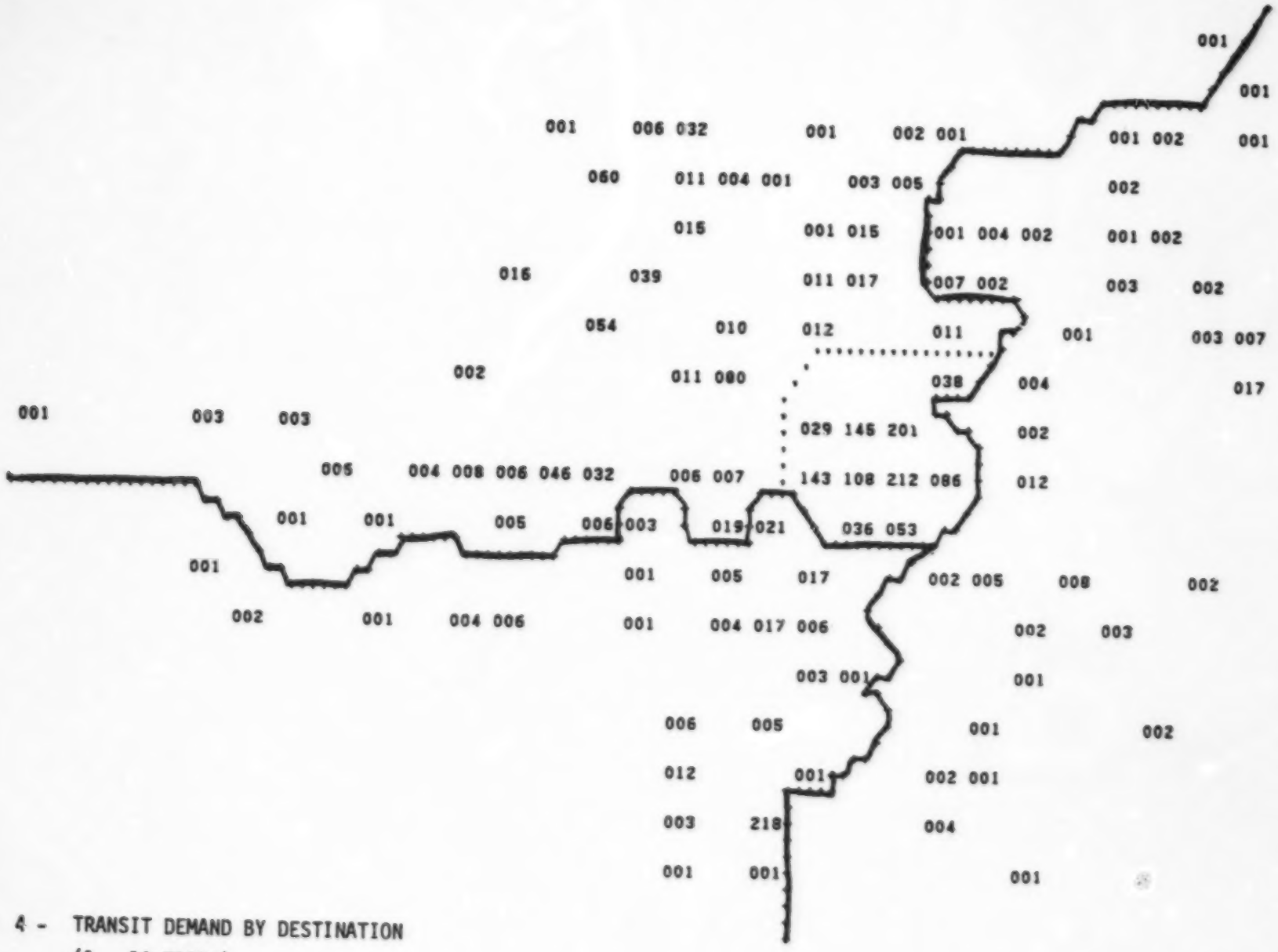

Figure 4 - TRANSIT DEMAND BY DESTINATION
(1 = 10 TRIPS)

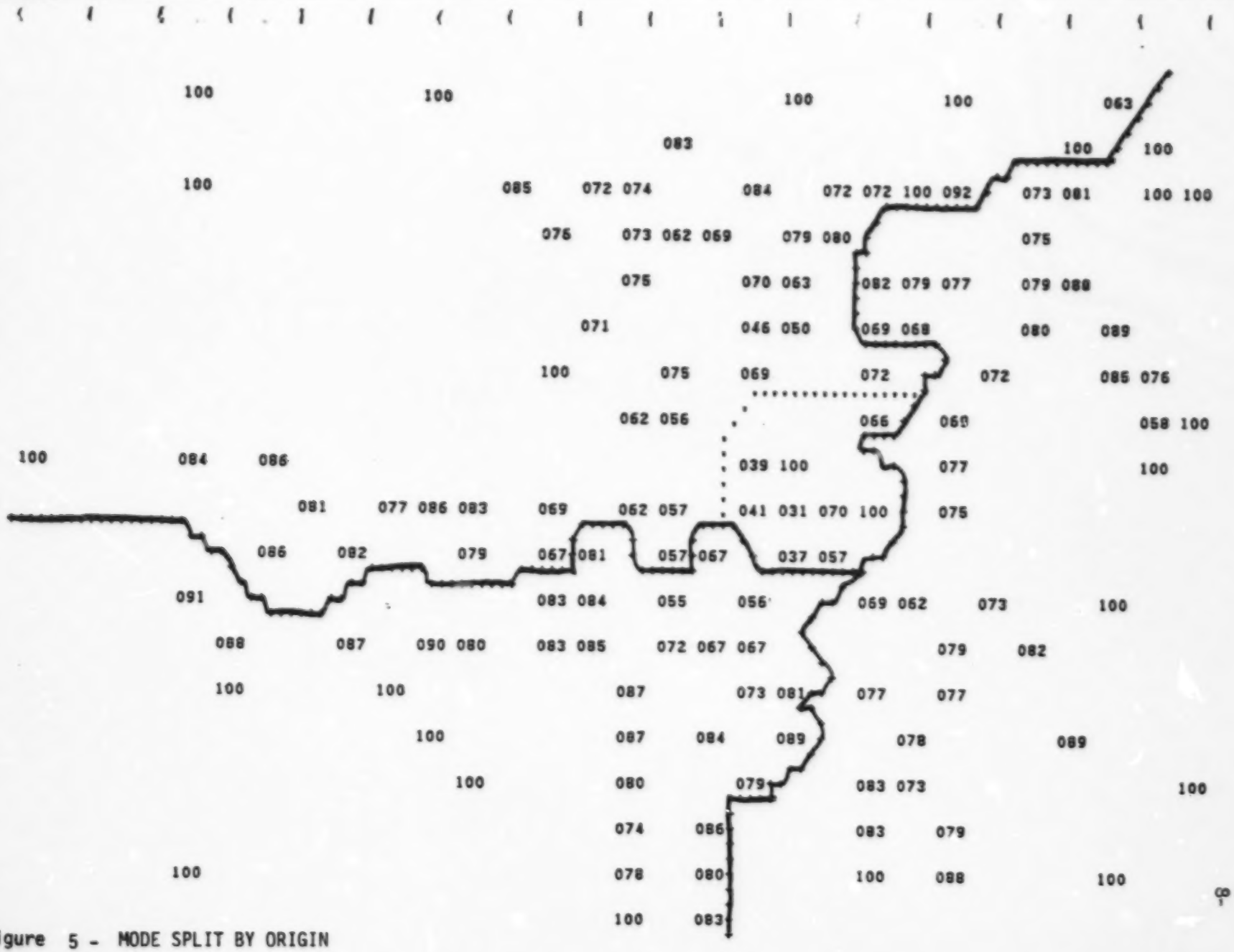

Figure 5 - MODE SPLIT BY ORIGIN

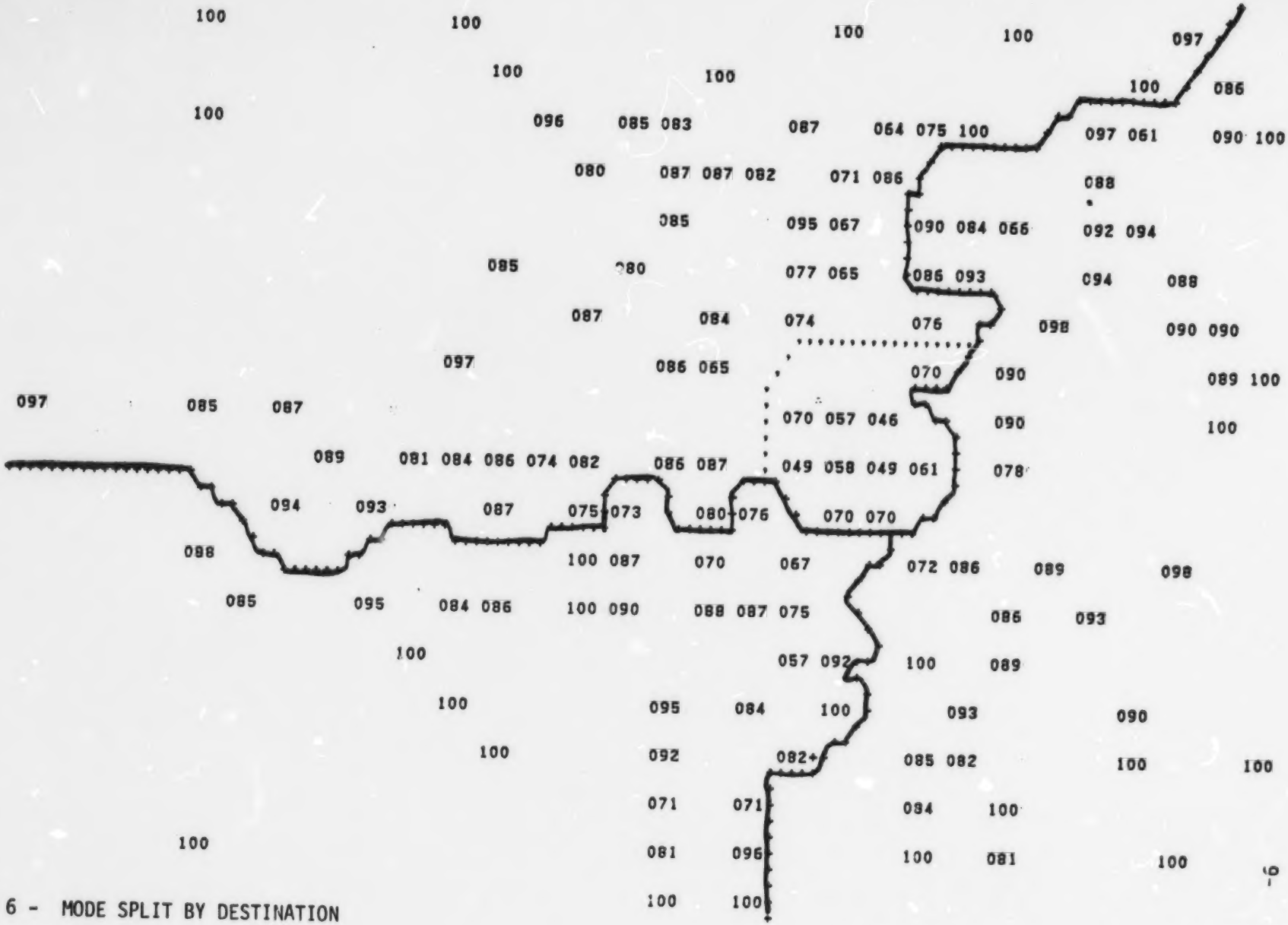

Figure 6 - MODE SPLIT BY DESTINATION

of this computation is a set of origin to destination travel times, that constitute the O/D matrix of

. road travel times (6)

and by tracing a set of shortest paths on the links that carry flow we obtain the O/D matrix of

. distance by road (7)

Then, the transit O/D Matrix (2) was used to calibrate the transit assignment. In essence, this calibration determines the coefficients of generalized time in the expression

$$IMP = WALK^* (\text{access} + \text{egress time}) + WAIT^* \sum_{\ell=1}^{n} w_\ell + \sum_{\ell=1}^{n} T_\ell + n^*WPEN$$

where

IMP is the travel time or impedance in generalized minutes

WALK is the weight of walking time (access-egress) relative to the in-vehicle time

WAIT is the weight of waiting time relative to the in-vehicle time

WPEN is a constant penalty added to the impedance for each wait

W_ℓ is the waiting time for the ℓth line and

T_ℓ is the in-vehicle time spent on the ℓth line, which is assumed to have a weight of 1.0 in the impedance calculations.

The best values obtained in trying to match observed transit volumes with predicted transit volumes were WALK = WAIT = 3 and WPEN = 0.

Thus, by applying these factors we obtained the O/D matrix of

. transit impedance (8)

and by tracing the shortest paths used we determined the O/D matrix of

. No. of transit transfers (9)

The Socio-Economic Variables

Since our approach is to calibrate a zonal aggregate modal split function, we extracted from the survey data the average car ownership/household/zone, the proportion of men travelling at the peak hour and the proportion of students travelling at the peak hour for each O/D pair.

The other socio-economic variables were obtained by the City of Winnipeg from various sources and transmitted to us. The Census Survey data served to estimate the average income/household/zone and the O/D survey served to estimate the number of jobs/zone. The parking costs/month/zone were evaluated by using 1976 data and the number of parking/spaces/job/zone were evaluated by using 1976 data as well. The transit fare is well known and the auto operating costs were estimated by considering an operating cost of $.20/mile.

Thus, in all, a file was constructed which consisted of the dependent variable, the observed modal split, and the independent explanatory variables described above. This file contained the records for the 1232 O/D pairs that had 20 or more trips in the expanded matrix (3) of trips by both modes. The main reason for considering only these O/D pairs is that the modal split for O/D pairs with smaller demand will indicate mostly the use of the car,

and thus bias the calibration of any modal split function. It is also worthwhile to mention that only 672 O/D pairs had observed transit demand; thus, even the 1232 selected O/D pairs contained many O/D pairs which had only observed auto demand.

The Functional Form of the Modal Split Function

The functional form chosen for the calibration is the logistic function. Although this form achieved recent fame in its use as a disaggregate probabilistic multiple choice function, we use it with aggregate data due to its ease of manipulation and its property of predicting choice values with a smooth type curve. The form that we used is

$$p_{au} = 1 \; / \; 1 + \exp\left(k_o + \sum_i k_i \, x_i\right)$$

where p_{au} is the proportion of trips that occur by auto

k_o is a constant and

k_i , i=1,...,m are the coefficients associated with the

k_i , i=1,...,n explanatory variables.

The explanatory variables x_i may describe an origin (such as car ownership), a destination (such as parking costs), and an origin/destination pair (such as the auto travel times and transit impedances).

A simple algebraic manipulation of the above equation results in the form

$$\ln\left(\frac{1-p_{au}}{p_{au}}\right) = k_o + \sum_i k_i \, x_i$$

which is used for calibrating k_o, k_i, i=1,...,n by multiple linear regression.

This method of estimation is often referred to as "Berkson-Theil" estimation, to acknowledge their early work [1], [7] with aggregate logistic function calibration.

Another functional form that we tried is the so-called "dogit" proposed recently by Gaudry [5] which adds to the logit form modal constants θ_{au}, θ_{tr} as follows:

$$p_{au} = \frac{1}{1 + \theta_{au} + \theta_{tr}} \left\{ \frac{1}{1 + \exp\left(k_o + \sum_i k_i x_i\right)} + \theta_{au} \right\}$$

However, in the trials that we performed, the best values for θ_{au} and θ_{tr} were always zero; the logistic function was satisfactory and, as the θ parameters represent captivity to travel modes, neither mode seemed to have a captive market.

The actual calibration of the modal split model spanned a period of eight months during which several hundred regressions were run. It is of little interest to report on all the modal split functions that were tried out, however we will report on those that seemed most acceptable for use in the equilibration procedure.

First, we calibrated the logit function on all the O/D pairs that has 20 or more trips by both modes in the expanded matrix of total trips. The coefficients of this function may be referenced in the first line of Table 1, as well as the coefficient of correlation R, R^2 and the true coefficient of correlation and its square, that are obtained by comparing the predicted and

observed modal shares. There were 1232 O/D pairs included in this calibration that included 400 O/D pairs that had no observed trips by transit. For the latter, we modified the data by indicating a modal share of .01 for transit in order to avoid the computation of log (0).

Next, we calibrated modal split functions by including O/D pairs that had an increasing number of total trips. The coefficients of the function obtained by including all the O/D pairs that have 60 or more trips by all modes are shown on the second line of Table 1. There are only 31 O/D pairs that have no observed transit trips out of a total of 242 O/D pairs included in this calibration. It is interesting to note that the observed modal share for the auto decreases from .7315 to .6804. This is due to the fact that there are less O/D pairs which have no observed transit trips. As may be expected the R^2 increases for this second model, while the same variables remain significant. Since the downtown oriented trips are often served better by transit and, due to other reasons as well, often exhibit different modal choice behaviour, we calibrated a modal split function for the trips oriented towards the downtown zones and another model for all other trips, by using the O/D pairs with 20 or more total trips. The results are shown in the third and fourth line of Table 1. Although the R^2 values for these functions are quite low the true R^2 is slightly higher than that obtained with the global model, that is .463 compared with .445.

We tested the modal split functions obtained in this way, by predicting the modal shares for the base year and then multiplying by the total number of trips, to obtain the trips taken by auto. The results of theses tests are shown in Table 2. The first three lines of this table present the absolute

NUMBER OF O/D PAIRS USED	AVERAGE OBSERVED MODAL SPLIT	MULTIPLE R	R^2	COEFFICIENTS (T-STATISTIC VALUES)						TRUE R^2	TRUE R
				PARKING COST	CAR OWNERSHIP	PROPORTION MEN	TRANSIT IMPEDANCE	AUTO TIME	CONSTANT		
20 +	.7315	.697	.4856	.0964 (13.935)	-2.665 (-11.024)	-2.361 (-11.622)	-.0189 (-6.549)	.0209 (3.022)	2.696 (11.166)	.445	.667
60 +	.6804	.776	.602	.0245 (6.890)	-1.904 (-4.562)	-3.229 (-7.711)	-.0122 (-2.459)	.022 (2.158)	2.563 (6.268)	.602	.776
20 + DT	.5642	.603	.364	.0934 (4.696	-2.423 (-4.929)	-2.512 (-7.178)	-.0131 (-1.776	.0147 (.985)	2.340 (4.710)	.463	.680
20 + ODT	.8244	.604	.365	.3632 (8.048)	-2.689 (-9.374)	-2.014 (-7.723)	-.0219 (-6.662)	.014 (1.724)	2.739 (9.146)		

Table 1 - The Modal Split Functions

NUMBER OF O/D PAIRS USED						
	TOTAL NUMBER OF TRIPS	ABSOLUTE ERROR	% ERROR	TOTAL NUMBER OF TRIPS	ABSOLUTE ERROR	% ERROR
20 +	19,905	346	1.74	35,065	2888	8.24
60 +	8.547	17	0.20	14,254	741	5.20
20 + DT ODT	19,905	542	2.72	35,065	2024	5.77
60 + / 20 +	19.905	-64	-0.32	35,065	1333	3.8

Table 2 - Errors introduced by the modal split functions in fit to base year data

Variable	Constant (k_o)	(Transit impedance)^2	(auto time)^2	(proportion men)^2	(parking cost)$^{\frac{1}{2}}$	ℓn (income)
Coefficient	16.566	-0.0000758	0.000242	-2.856	.360	-1.752
(T-statistic)	(4.97)	(-2.89)	(1.13)	(-9.05)	(8.46)	(-4.97)
R^2	.64					
R	.80					

Table 3 - Modal split function with transformation of variables

and % errors when the models are applied to the same O/D pairs that were used in their calibration. The fourth line of the table presents the absolute and % error when the model obtained by using only the O/D pairs with 60 or more trips is applied to the O/D pairs with 20 or more trips. It is clear that the best predictions for the base year are obtained with the latter procedure and this is due to the bias introduced in the models calibrated with 20 or more observations by the numerous O/D pairs that have no observed transit usage. This analysis would seem to point out that heteroschedasticity would be present in the models calibrated with 20 or more observations, however we did not pursue this matter, since we did not use these models for equilibration.

We proceeded next to apply transformations on the independent variables and the best modal split function obtained in this way by including the 242 O/D pairs with 60 or more trips is shown in Table 3. The R^2 value is increased slightly from .60 to .64, which is not significant enough to justify its use.

Another approach that we tried was to subdivide the origins into subgroups by using a criterion related to the error introduced by the modal split function, when it is used for predicting the O/D matrices for the base year. We reasoned that errors on individual O/D pairs were unavoidable, but that the model should not distort the O/D matrix, that is, there should not be too much bias introduced on demand totals by origins and destinations. Thus, we subdivided the origins into subgroups according to the error introduced by the model on origin totals; origins which had negative deviations were grouped together and similarly origins with positive errors and origins

which had acceptable errors formed two other subgroups. Finally we obtained four modal split models which are shown in Table 4. Inspection of these models reveals that the disaggregation has the effect of reducing the significance of some of the explanatory variables due to the relatively small number of O/D pairs in each subgroup.

Finally, we would like to point out that all the modal split functions that we calibrated share the property of being relatively "flat" with respect to the auto travel times and bus impedances. In Figure 7 and 8 we illustrate this with a plot of the modal split function, obtained with the O/D pairs with 60 or more trips, against the auto travel time and the bus impedance. The implication of this property is that fairly major changes in travel times are necessary to induce significant changes in the modal split.

Conclusions

In summary, the calibration of a zonal aggregate modal split function is a non trivial task, even when using relatively simple estimation procedures (from the econometric viewpoint). In spite of the difficulties encountered, both ourselves and the staff of the City of Winnipeg are satisfied with the results obtained and consider that we have an adequate model for use for transportation planning purposes. The model has been integrated in the urban development plan review process and is being used by the City of Winnipeg for short term planning purposes.

Variables	Model 3a		Model 3b		Model 3c		Model 3d	
Constant	2.352	(1.46)	1.516	(2.46)	3.071	(2.03)	2.935	(5.52)
transit impedance	-0.0133	(-.70)	-0.0101	(-1.68)	-0.0315	(-3.27)	-0.0139	(-1.04)
auto time	0.0334	(.83)	0.0253	(1.99)	0.0733	(3.84)	0.0323	(1.05)
proportion men	-2.799	(1.365)	-3.719	(-6.64)	-3.101	(-3.26)	-2.499	(-3.32)
parking cost	.0959	(2.73)	.0968	(7.00)	.0471	(2.27)	.0332	(1.47)
car availability	-3.023	(-1.89)	-1.232	(-2.10)	-1.502	(-.96)	-1.844	(-2.88)
R^2	.72		.67		.78		.64	
R	.85		.82		.88		.80	
No. of O/D pairs	26		135		36		45	

The number in parentheses is the T-statistic associated with the parameter.

Table 4 - Modal split models by subgroups

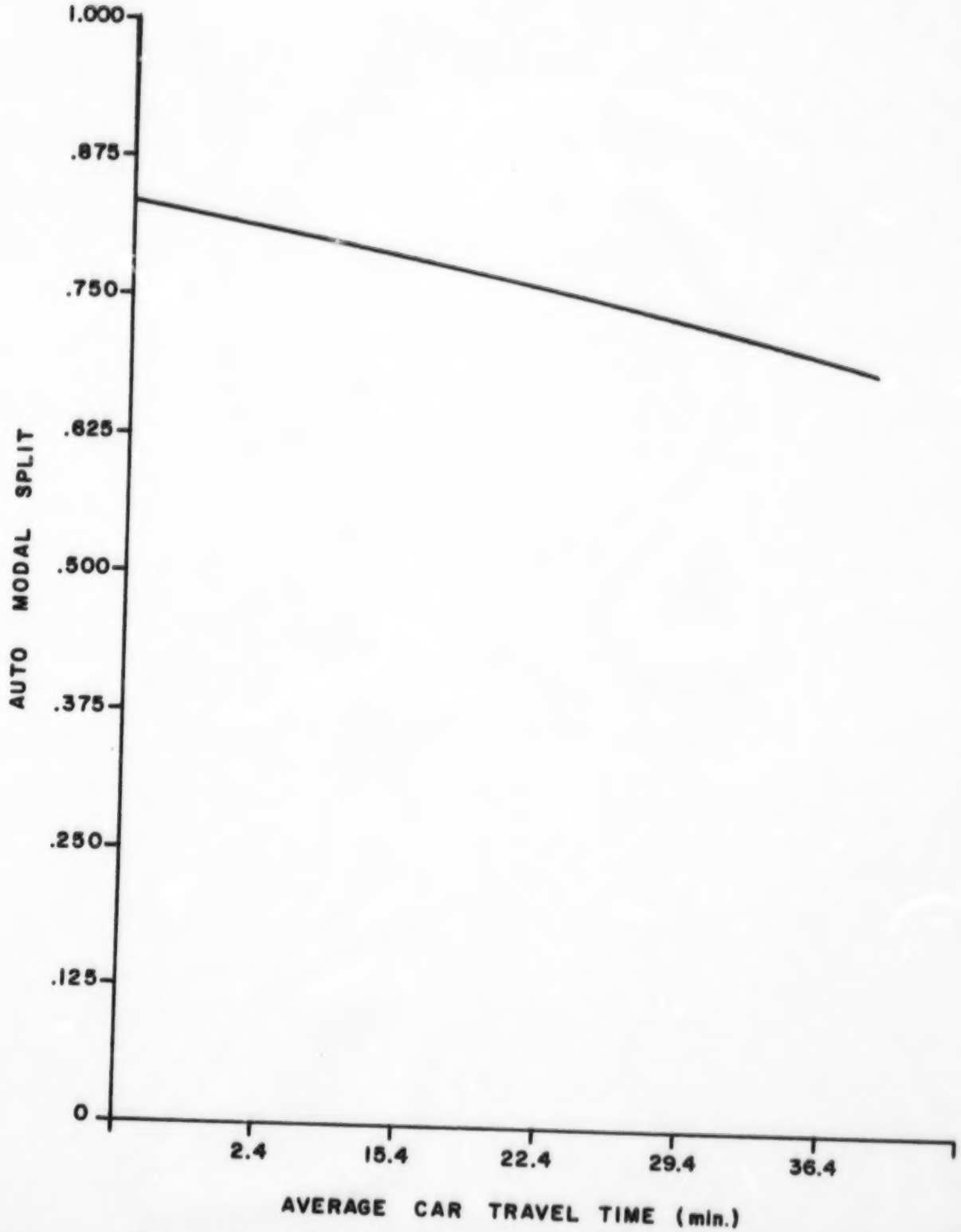

Fig. 7 – THE MODAL SPLIT FUNCTION (MODEL I)

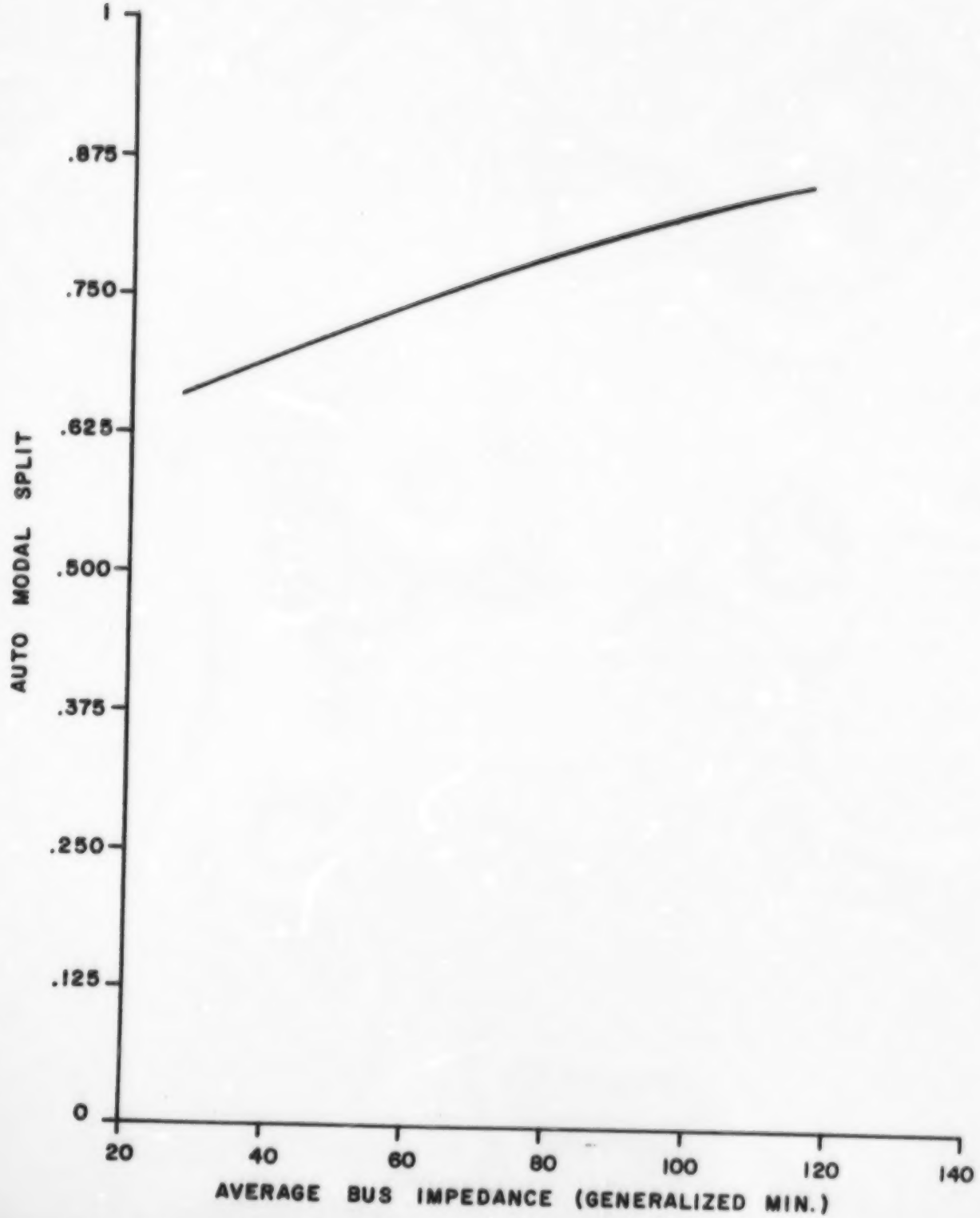

Fig. 8 – THE MODAL SPLIT FUNCTION (MODEL I)

ACKNOWLEDGMENTS

We are indebted to Ed Guertin from the transportation planning staff of the City of Winnipeg for preparing some of the data required for this work and to Claude Achim and Stefano Pallottino for assisting with some of the computations.

REFERENCES

[1] Berkson, J., "Maximum Likelihood and Minimum x^2 Estimate of the Logistic Function", Journal of American Statistical Association, 50, pp. 130-161, 1955.

[2] Florian, M., "A Traffic Equilibrium Model of Travel by Car and Public Transit Modes", Transportation Science, Vol. 11, No. 2, pp. 166-179, 1977.

[3] Florian, M., Chapleau, R., Nguyen, S., Achim, C., James, L. and Lefebvre, J., "EMME - A Planning Method for Multi-Modal Urban Transportation Systems", Proceedings of the Third World Conference on Transport Research, pp. 512-519, Edited by E.J. Visser, Martimes Nijhoff, The Hague, 1977.

[4] Florian, M., "EMME - Summary Report", publication #101, Centre de recherche sur les transports, Université de Montréal, 133p., 1979.

[5] Gaudry, M. and Dagenais, M., "The Dogit Model", publication #82, Centre de recherche sur les transports, Université de Montréal, 14p., 1978. To appear in Transportation Research, 1979.

[6] Nguyen, S., "Equilibrium Traffic Assignment Procedures with Elastic Demand", publication #32, Centre de recherche sur les transports, Université de Montréal, 48p., 1977.

[7] Theil, H., Principles of Econometrics, pp. 65-76, Wiley, New York, 1971.